METHOD OF FORMATTING A DATA FLOW BY CODING BASED ON THE SEQUENCE OBJECTS OF ANIMATED IMAGES

Ins a'

Related Art

The present invention relates to a method of processing a
data stream for object-based coding of moving image
5   sequences which may have any size and shape.

Ins a²

MPEG-4 Video Verification Model Version 7.0, Bristol, April
1997, MPEG-97/N 1642, ISO/IEC JTC/SC 29/WG 11 specifies an
encoder and decoder for object-based coding of moving image
10   sequences, where rectangular images of a fixed size are no
longer coded and transmitted to the receiver within a video
session (VS), but instead video objects (VO) of any size and
shape are coded and transmitted. These video objects may
then be further subdivided into different video object
15   layers (VOL) to represent different resolution levels of a
video object, for example. The image of a VO of a certain
layer in the plane of the camera image at a certain time is
the video object plane (VOP). Thus, the relationship between
VO and VOP is equivalent to the relationship between image
20   sequence and image in transmission of rectangular images of
a fixed size.

The syntax for transmission of a VOP specifies first the
signaling of the local time base of a VOP. This indicates
25   the time with respect to previously transmitted VOPs at
which the instantaneous VOP is to be displayed. Figure 1
shows the syntax structure for elements VS, VO, VOL and the
relevant parts for element VOP.

30   The parts of the VOP syntax shown here are relevant in this

connection. The "modulo time base" element indicates the local time base of the VOP in increments of 1000 milliseconds, and the "VOP time increment" element also indicates the local time base in increments of one

5   millisecond. The "VOP prediction type" element indicates which type of prediction is to be used for the VOP. There are four possibilities here: I-VOP, i.e. no prediction is used, P-VOP, i.e. the prediction is based on the preceding VOP, B-VOP, i.e. the prediction is based on the preceding

10   and following VOPs, and S-VOP where the prediction is based on a SPRITE-VOP which is either transmitted once at the start of the video session or is derived from the reconstructed data during transmission.

15   In addition to transmission of the local time base of a VOP, the syntax specifies a possibility of signaling the "coded/not coded" state for a VOP. In the case of the "not coded" state for the VOP, no additional data is transmitted after the corresponding signaling elements, and if there is

20   a new VOP, transmission thereof is begun. On the receiver end, a "not coded" VOP is not decoded further and is not displayed.

Here the "video object layer shape" element, which is

25   specified in the area of the header info of the syntax of the respective VOL, indicates whether the VO is a rectangular VO (== 0) or is a VO of any size and shape ($\neq$ 0). Then for the case of a VO of any size and shape, the width of the rectangle surrounding the VOP is indicated with

30   the help of the "VOP width" element. If this width is set to the value 0, this signals that the VOP has the "not coded" state. Then the transmission of the data of the instantaneous VOP is terminated and transmission of the next

2

VOP is begun.

~~Advantages of the Invention~~

5   With the ~~measures~~ method according to the present invention, it is
     possible to transmit less data for a non-coded video object,
     i.e for a video object that is not to be displayed
     immediately. In contrast with the aforementioned related
     art, it is simpler and more comprehensible to use a definite
10   element for signaling the state of whether or not a video
     object is to be displayed.

     With the method according to the present invention, it is
     also possible to transmit and thus to signal the coded/not
15   coded state for rectangular VOs, which had not been possible
     with the implementation according to the related art.

     The signaling information indicating whether a video object
     is coded or not coded may be inserted before or after the
20   local time base information in the data stream. If the
     signaling information is inserted before the local time base
     information, even less data need be transmitted for a non-
     coded VOP than when the signaling information is inserted
     after the local time base, because in this case the local
25   time base information is not transmitted. However, in this
     case, the "blanking out," i.e. suppression of the display of
     a video object, is no longer possible at a very specific
     point in time, but instead it can only take place at the
     next time following the receipt of the non-coded VOP, when
30   an image is displayed at the receiver end.

3

Description of Embodiments

~~Diagram~~ *Figure* 2 shows the structure of the data stream for the
transmission of video objects. At the beginning (first line
of the diagram) the "video session start code" element is
transmitted and then the information for video objects 1, 2,
..., n is transmitted. At the end, the "video session end
code" appears. The second line shows the structure of the
transmission format for "video object 1." It begins with the
"video object start code" followed by the "video object
identification" and the elements for "video object layers" 1
through n. The third line shows the structure of a single
"video object layer" element. It begins with the "video
object layer start code" followed by the "video object layer
identification," the "header info" and elements 1 through n
for the "video object plane." The fourth line shows the
structure of a single "video object plane" element. It
begins with the "VOP start code" followed by the local time
base information "modulo time base" and the "modulo time
increment" element. This structure thus corresponds to the
structure according to ~~Diagram~~ *Figure* 1. In contrast to ~~Diagram~~ *Figure* 1,
however, a new element in the form of signaling information
is always inserted into the data stream according to the
present invention, indicating whether the video object is to
be decoded for playback or displayed. The signaling
information is also inserted regardless of the external form
of a video object. This signaling information is composed of
the "VOP coded" element and is defined so that the value 0
denotes the "not coded" state and the value 1 denotes the
"coded" state. For the receiver, it is necessary to define
the fact that the corresponding VO is no longer displayed
for the case "VOP coded == 0" at the time indicated by the
local time base or at the next following time when an image

4

is displayed at the receiver end. In contrast with the
implementation according to ~~Diagram~~ *Figure* 1, there is no longer
any signaling by the "VOP width" element.

5   The "VOP coded" element can also be inserted into the data
stream after the "VOP prediction type" element. ~~Diagram~~ *Figure* 3
shows another embodiment of the present invention. "VOP
coded" signaling information is then placed directly after
the "VOP start code" element, i.e before the local time base
10   information "modulo time base." For this embodiment as well,
"VOP width" signaling is no longer performed. In contrast
with the first embodiment (~~Diagram~~ *Figure* 2), even less data need
be transmitted for a non-coded VOP, because the local time
base need not be transmitted. However, in this case the
15   "blanking out," i.e. no longer displaying a VO, is no longer
possible at a very specific point in time, but instead it
can only take place at the next time following the receipt
of the "non-coded" VOP, when an image is displayed at the
receiver end.

20